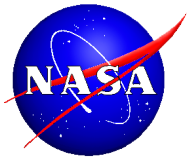


Jonathan Gleason

DATA MANAGEMENT STATUS

Data Management Team
CERES Science Team Meeting
Geophysical Fluid Dynamics Laboratory
October 22, 2012

Overview

- DMT Structure & Function
- DMT Activity Since Last Meeting
- Platform Migration Status
- Data Availability: Current and Planned
- FM5: Processing Status
- Processing Automation (CATALYST)

Data Management Team Members

Task Manager:
Walt Miller

Systems/ Optimization:
Nelson Hillyer
Josh Wilkins

Production / Optimization:
Lisa Coleman
Carla Grune

Instrument:
Denise Cooper
Thomas Grepiotis
Mark Timcoe
Dianne Snyder

ERBElike:
Dale Walikainen
Jeremy Lande

Configuration Management:
Tammy Ayers
Joanne Saunders

Clouds:
Sunny Sun-Mack
Ricky Brown
Yan Chen
Liz Heckert
Rita Smith
Sharon Gibson

Convolution:
Igor Antropov

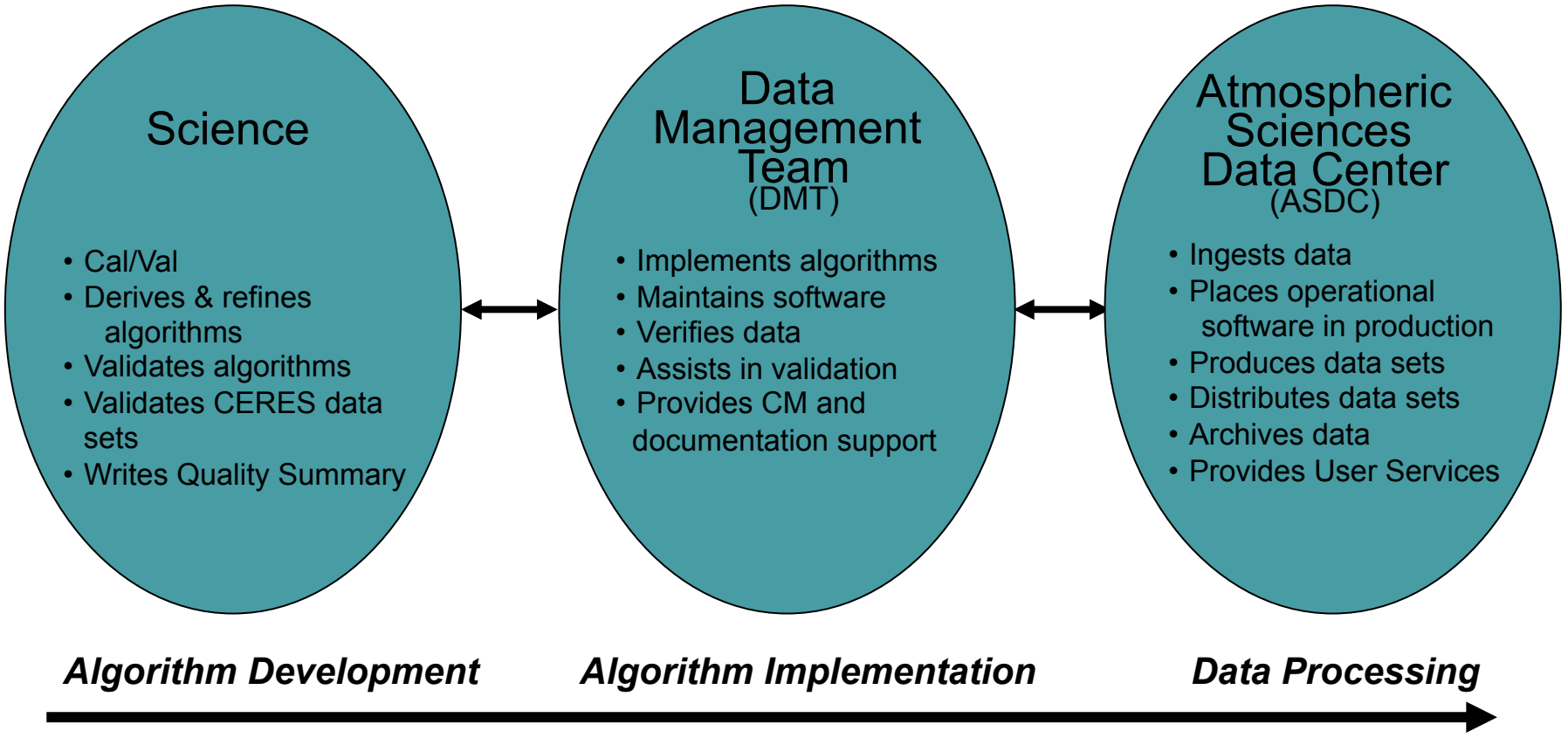
Inversion:
Victor Sothcott

SARB:
Tom Caldwell

TISA Gridding:
Raja Raju
Forest Wren

TISA Averaging:
Cathy Nguyen
Dennis Keyes

CERES Organization



CERES Production Software Overview

Subsystem Number	Subsystem Name	Number of PGEs	LOC (to nearest 1K)	Publicly Available Data Products	Product Frequency
1	Instrument/Pre-Processor	1	5K		
1	Instrument	6 + lib	156K	BDS	1/day
2	ERBE-like/ Inversion	5 + lib	31K	ES-8	1/day
3	ERBE-like/ TSA	2	14K	ES-9, ES-4	1/month
4.1 – 4.4	Clouds/VIIRS Subset Code	1	21K		12/hour
4.1 – 4.4	Clouds	10	351K		
4.5 – 4.6	Inversion	13	147K	SSF	1/hour
5	SARB	6	166K	CRS	1/hour
6 & 9	TISA-Gridding	13	56K	SSF1deg-Hour, ISCCP-D2like-Day/Nit	60/month, 36/month, 1/month
11	GGEO	1	6K	ISCCP-D2like-GEO	1/month
7.2	Synoptic SARB	1	47K		
7.1 & 8 10	TISA-Averaging	6	184K	SSF1deg-Day, SSF1deg-Month, SYN1deg-(3Hour, M3hour, Mhour, Month)	1/day, 1/ month, 1/ month 5/month
12	MOA	2	14K		
	CERESlib		120K		
	Total	**67**	**1,318K**		

Activity since Spring STM

32 deliveries since 5/1/2012

- **Instrument (5)**
  - NPP planned post launch delivery
  - 3 NPP planned NPP deltas
  - 1 update due to IDPS Leap Second implementation
- **ERBE-Like (4)**
  - Delivery of all PGEs to x86 platform
  - 3 sample read software packages
- **Clouds (2)**
  - Edition 4 Beta FINAL delivery
  - Delta modification of monthly script
- **Inversion (2)**
  - Edition 4 Beta delivery
  - Edition 3 PGEs migrated to AMI-P
- **Tisa Averaging (3)**
  - 2 ancillary input files (SOURCE)
  - SYN1deg sample read SW
- **Tisa Grid (3)**
  - 3 ISCCP-D2like sample read SW
- **Inst SARB (2)**
  - Preprocessor migrated to AMI-P
  - 1 ancillary input files (MATCH)
- **GGEO (1)**
  - Delta for GGEOW (now offline)
- **CERESlib (2)**
  - 1 CATALYST support
  - 1 compiler & TK upgrade support
- **Perl_Lib (4)**
  - 3 CATALYST functionality deliveries
  - 1 Clouds Ed4 support
- **CATALYST/PR Database (2)**
  - PR Database Build 1
  - CATALYST PRDB login module

Computing Platform Migration

- Magneto -> AMI-P migration complete May 2013

Magneto (P4)		AMI-P (P6 & x86)
• Snow/Ice 1/16th mesh PGE (pending data source)	On Hold	• Instrument • ERBE-Like • TISA Averaging Ed3 • Synoptic SARB Ed3 • MOA Ed4 • ISCCP-D2like Merge • FLASHflux • Inversion Ed 3 • MOA G5.1 & G5.4 • Clouds Ed4 beta • Inversion Ed4 beta
• Clouds Ed1-CV	Feb 22	
• Inversion Ed1-CV	Nov 16	
• TISA Grid Ed3	Oct 26	
• ISCCP-D2like-day/nit	Dec 14	
• ISCCP-D2like-GEO	Dec 14	**In Development** • TISA Gridding Ed4 beta
• ~~Inst. SARB Ed2~~		

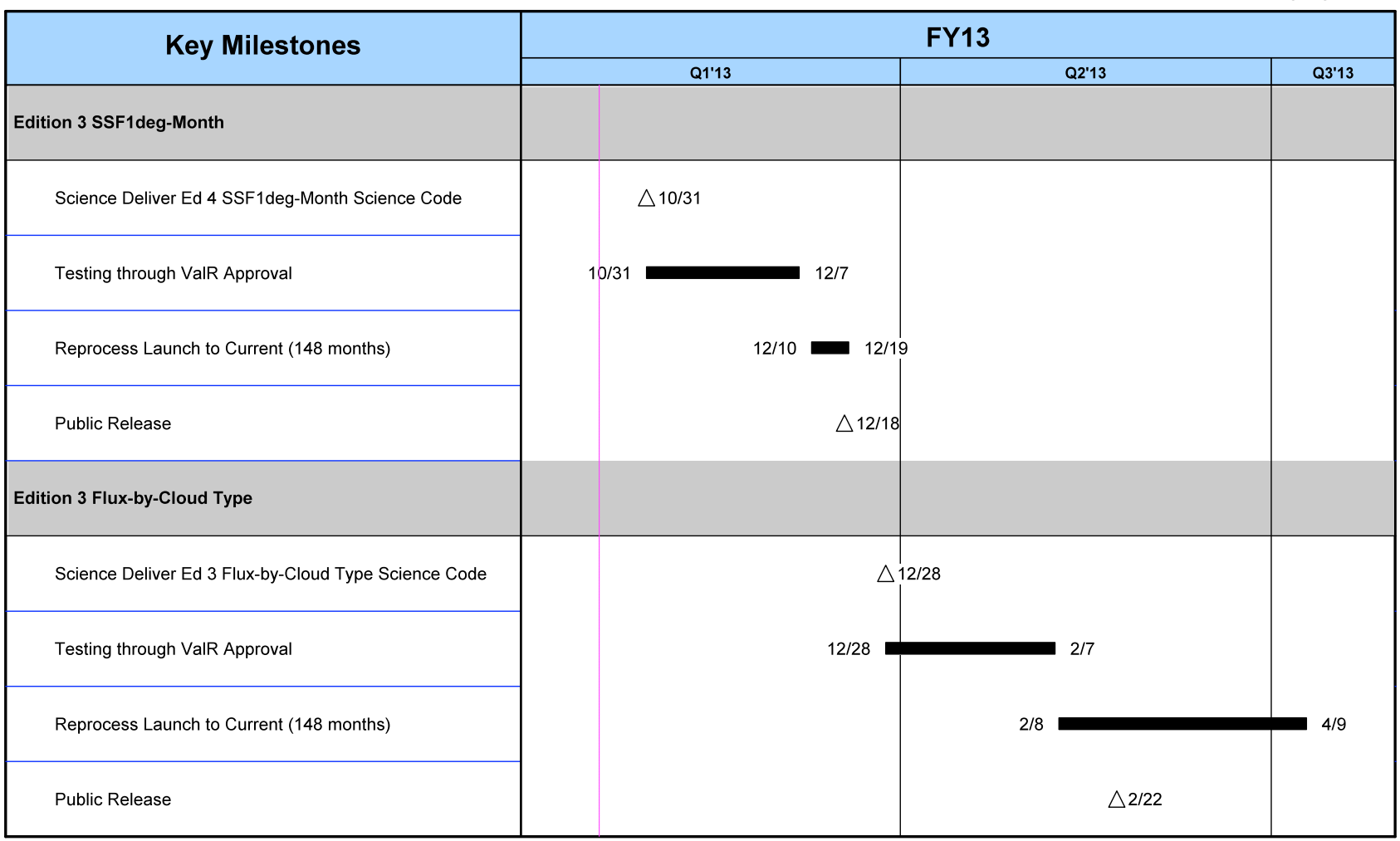
Current Data Availability

Product	Instruments	Available through	Publically Available
BDS (Ed 3)	Terra Aqua	Jan 1, 2012 Jan 1, 2012	Yes
SSF (Ed 3)	Terra Aqua	Dec 31, 2011 Dec 31, 2011	Yes
SFC (Ed 3)	Terra & Aqua	Nov 2011	Yes
SYN1deg (Ed 3)	Merged	Nov 2011	Yes
SSF Edition 4 Beta	Terra Aqua	Feb 6, 2002 Nov 17, 2003	No
SSF Edition 1-CV	Terra Aqua	Aug 23, 2012 July 31, 2012	No
ISCCP-D2like-Day ISCCP-D2like-Nite		Mar 2000 – June 2010	
ISCCP-D2like-GEO		Mar 2000 – Feb 2010	
ISCCP-D2like-Mrg		July 2002 – June 2010	

Edition 3 Planned Processing

CERES Science Software Delivery Summary Schedule
Edition 3 Deliveries

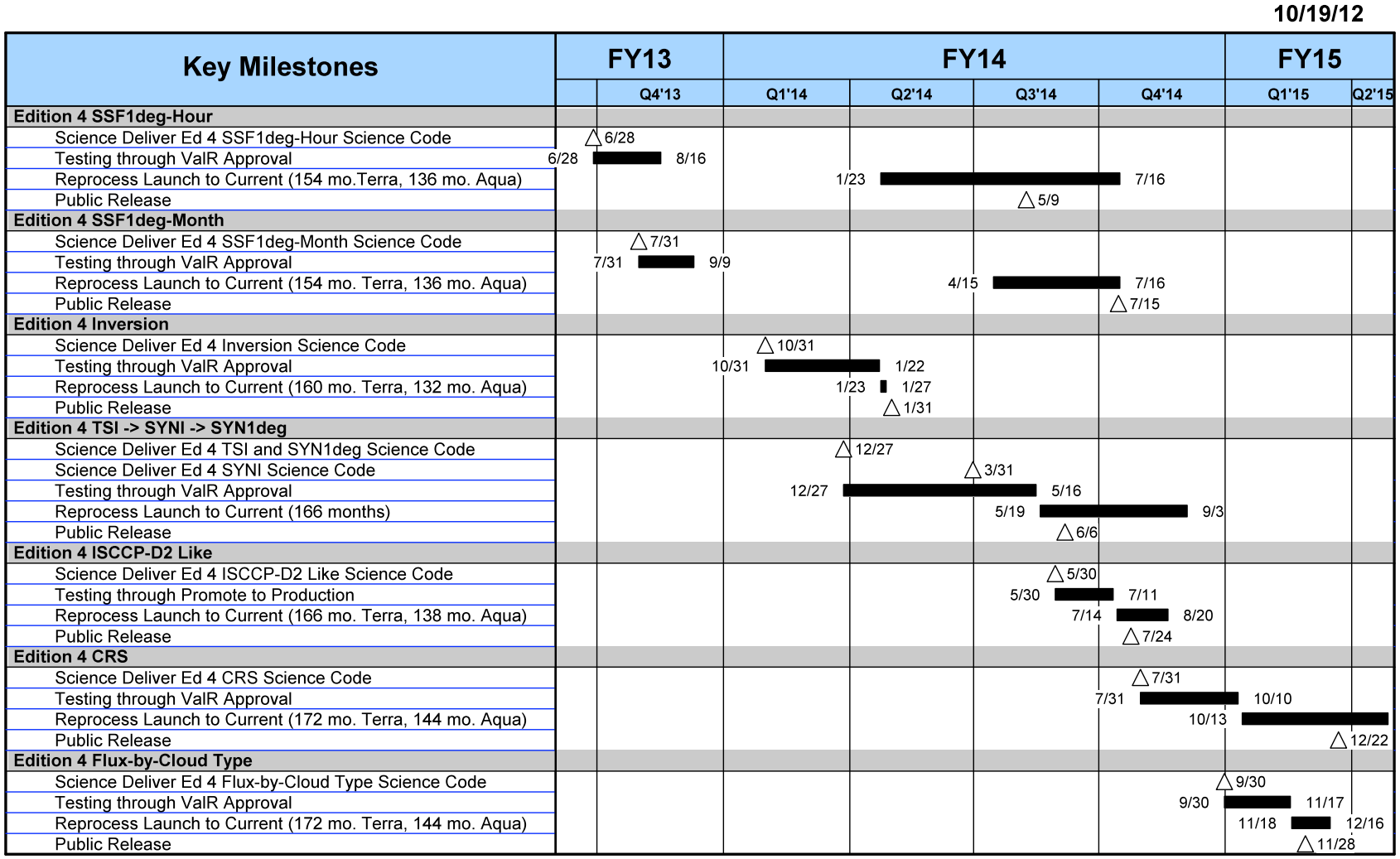
10/19/12



Edition 4 Planned Processing

CERES Science Software Delivery Summary Schedule
Edition 4 Deliveries

10/19/12



Edition 3 & 4 Planned Milestones

Product	Science Delivery to DMT	Target Public Release
Ed3 SSF1deg-Month	Oct 31, 2013	Dec 28, 2013
Ed3 Flux-By-Cloud-Type	Dec 28, 2013	Feb 22, 2014
Ed4 SSF1deg-Hour	June 28, 2013	May 9, 2014
Ed4 SSF1deg-Month	July 28, 2013	July 15, 2014
Ed4 Inversion	Oct 31, 2013 (ADM & SOFA groups)	Jan 31, 2014
Ed4 TSI & SYN1deg	Dec 27, 2013	June 6, 2014
Ed4 SYNI	Mar 31, 2014	
Ed4 ISCCP-D2like	May 30, 2014	July 24, 2014
Ed4 CRS	July 31, 2014	Dec 22, 2014
Ed4 Flux-By-Cloud-Type	Sept 30, 2014	Nov 28, 2014

NPP Data Processing Status

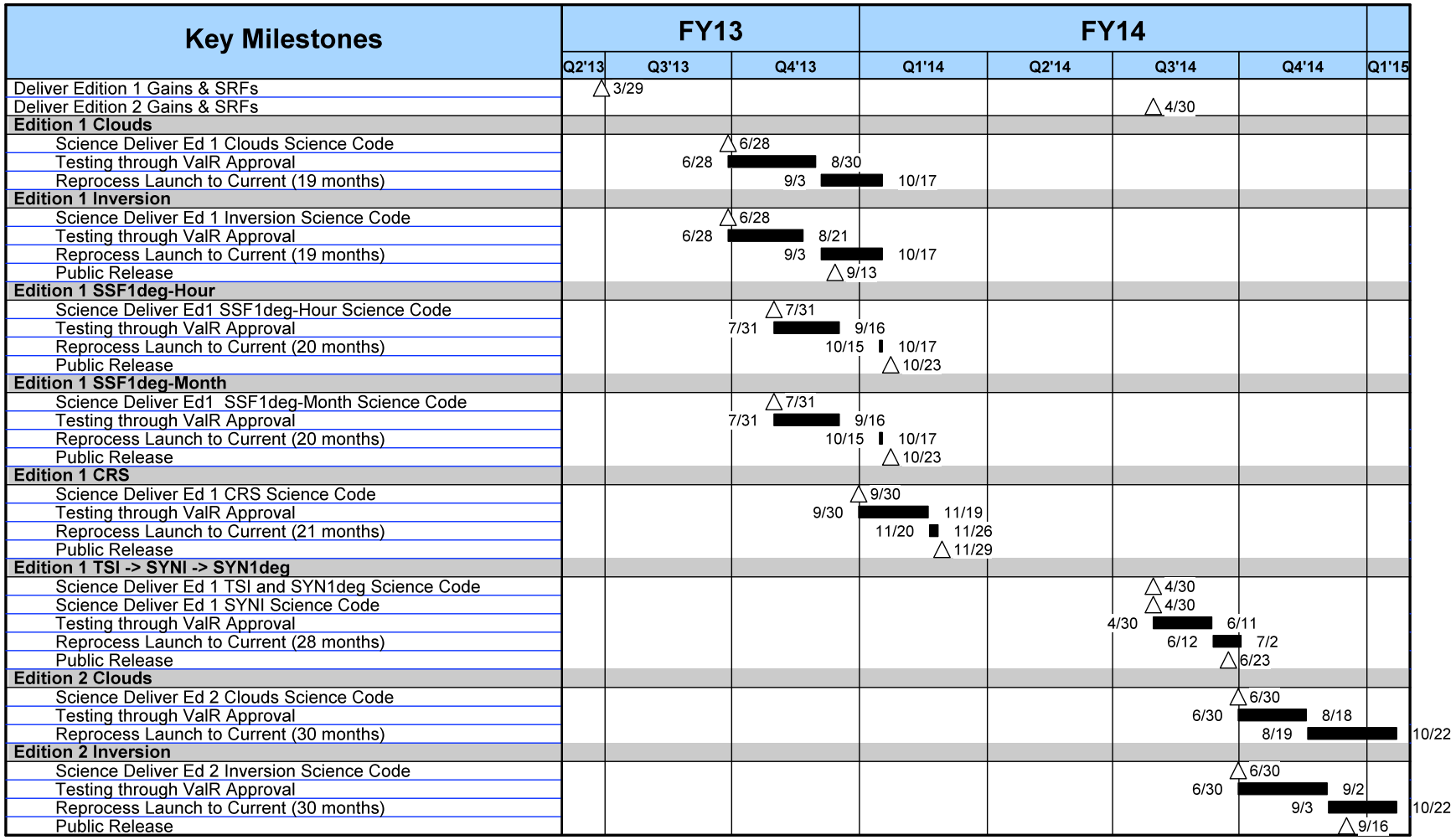
- Science data collection began Jan 27, 2012
- Forward processing stream Instrument (BDS) and ERBElike only
- First reprocessing with constant calibration occurred Aug 3rd-5th and data publically released Sept 18th
- Planned For Public Release in 2013
  - Edition 1 SSF
  - Edition 1 SSF1deg-Hour & SSF1deg-Month
  - Edition 1 CRS

NPP Planned Processing

CERES Science Software Delivery Summary Schedule

NPP Deliveries

10/19/12



NPP Planned Milestones

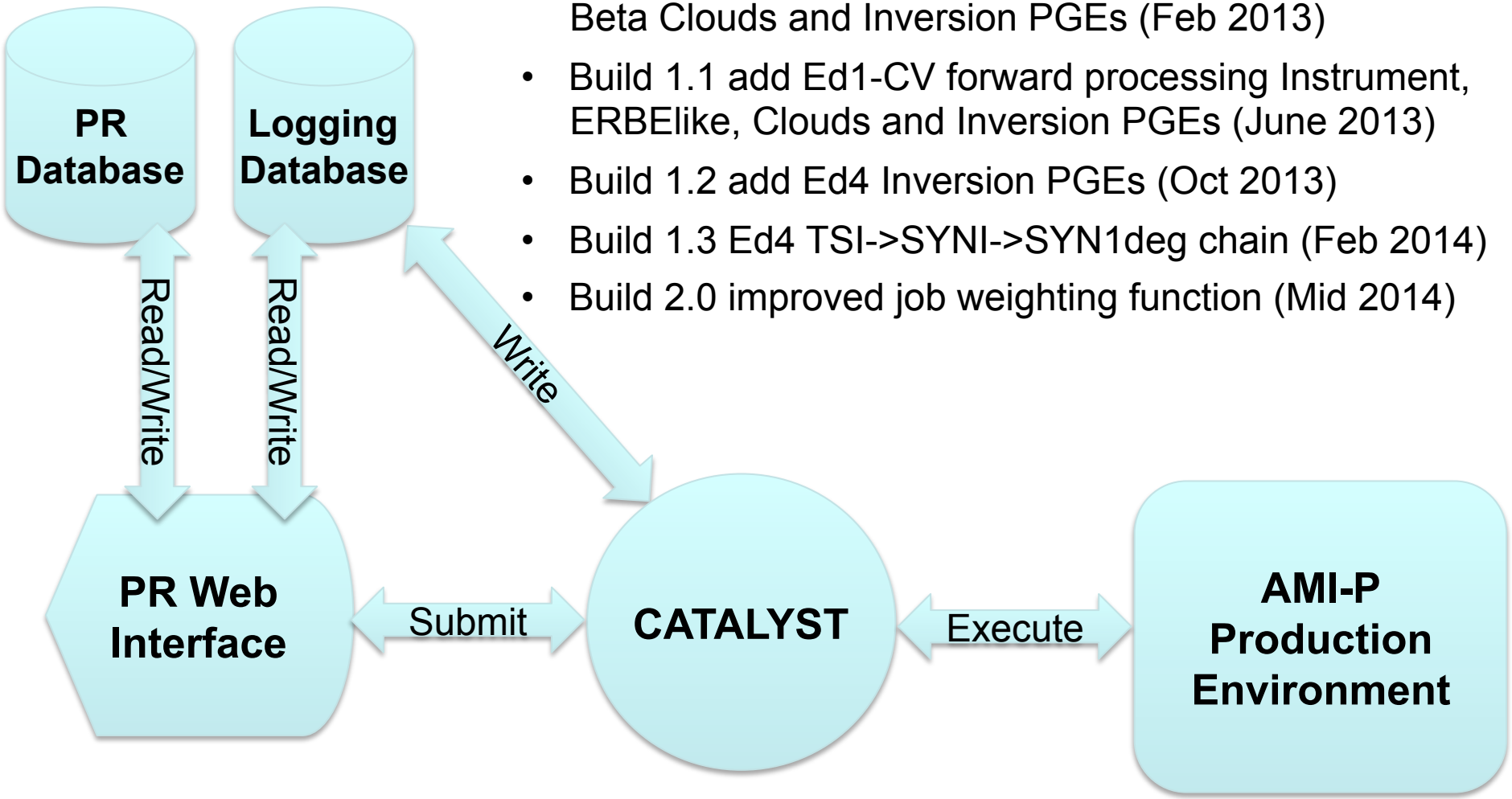
Product	Science Delivery to DMT	Target Public Release
Ed1 Clouds	June 28, 2013	Sept 13, 2013
Ed1 Inversion	June 28, 2013 (ADM & SOFA groups)	
Ed1 SSF1deg Hour & Month	July 31, 2013	Oct 23, 2013
Ed1 CRS	Sept 30, 2013	Nov 29, 2013
Ed1 TSI & SYN1deg	April 30, 2014	June 23, 2014
Ed1 SYNI	April 30, 2014	
Ed2 Clouds	June 30, 2014	Sept 16, 2014
Ed2 Inversion	June 30, 2014 (ADM & SOFA Groups)	

CERES Production Automation

- **C**ERES **A**u**T**om**A**ted job **L**oading **SYST**em (CATALYST)
- Goal
  - Maximize cluster throughput (we have 330 processors!)
  - Allow for graceful operator intervention or enhanced control
  - Accommodate regular PGE modifications
  - Function within existing scripting & grid engine framework
- Approach
  - Leverage existing PR DB effort to transmit production requests to automated system
  - System builds jobs from PR, tracks, & submits to cluster
  - Jobs submitted to grid engine using existing scripting
  - All job information logged in dedicated log DB
  - Log DB provides search capability for predecessor job status

CATALYST Architecture

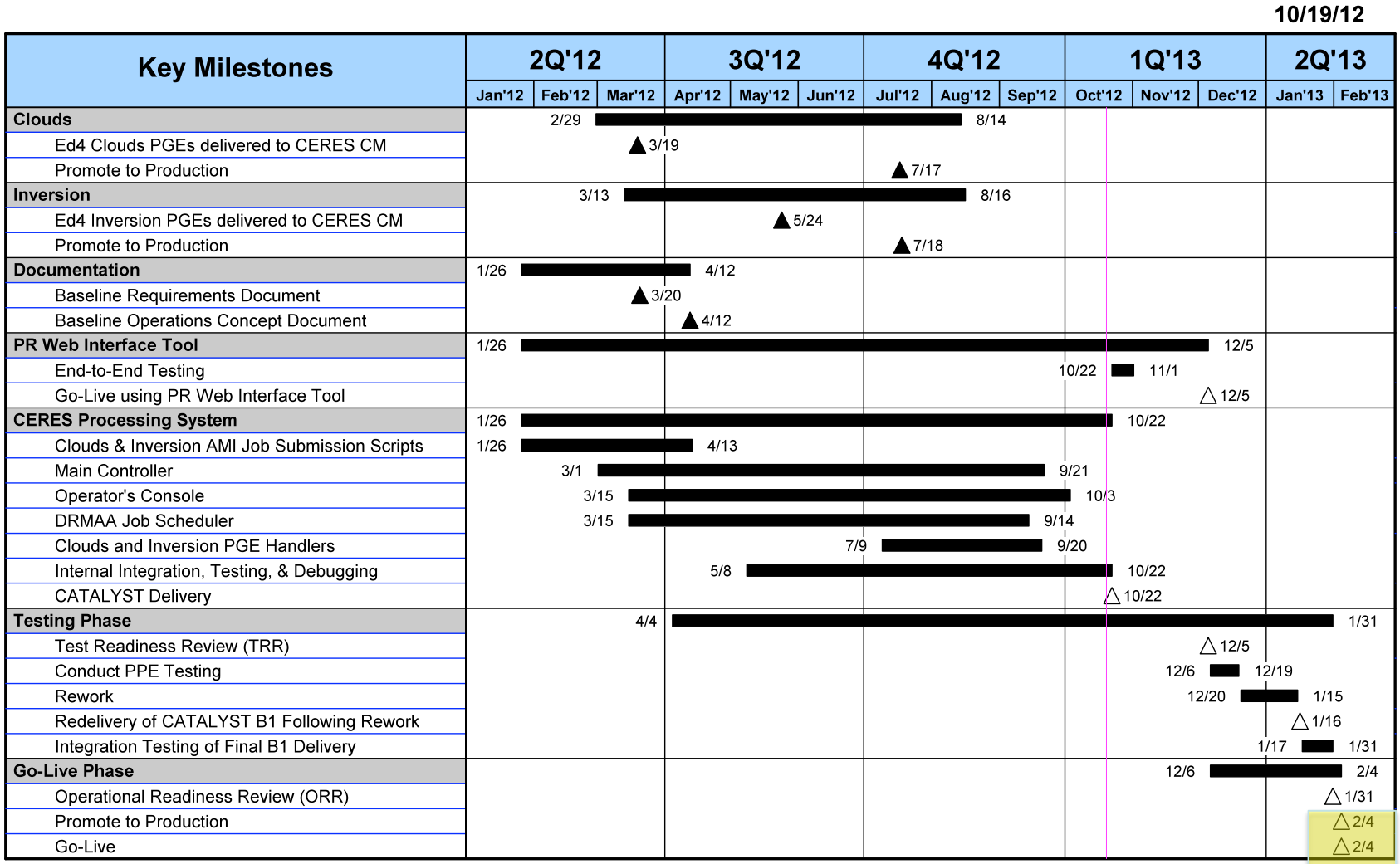
- Intensive development work started post NPP Launch
- Build 1 implements core functionality and runs Ed4 Beta Clouds and Inversion PGEs (Feb 2013)
- Build 1.1 add Ed1-CV forward processing Instrument, ERBElike, Clouds and Inversion PGEs (June 2013)
- Build 1.2 add Ed4 Inversion PGEs (Oct 2013)
- Build 1.3 Ed4 TSI->SYNI->SYN1deg chain (Feb 2014)
- Build 2.0 improved job weighting function (Mid 2014)



CATALYST Build 1 Schedule

CERES AuTomAted Loading sYSTem (CATALYST) Schedule
Summary View

10/19/12



Backup Charts

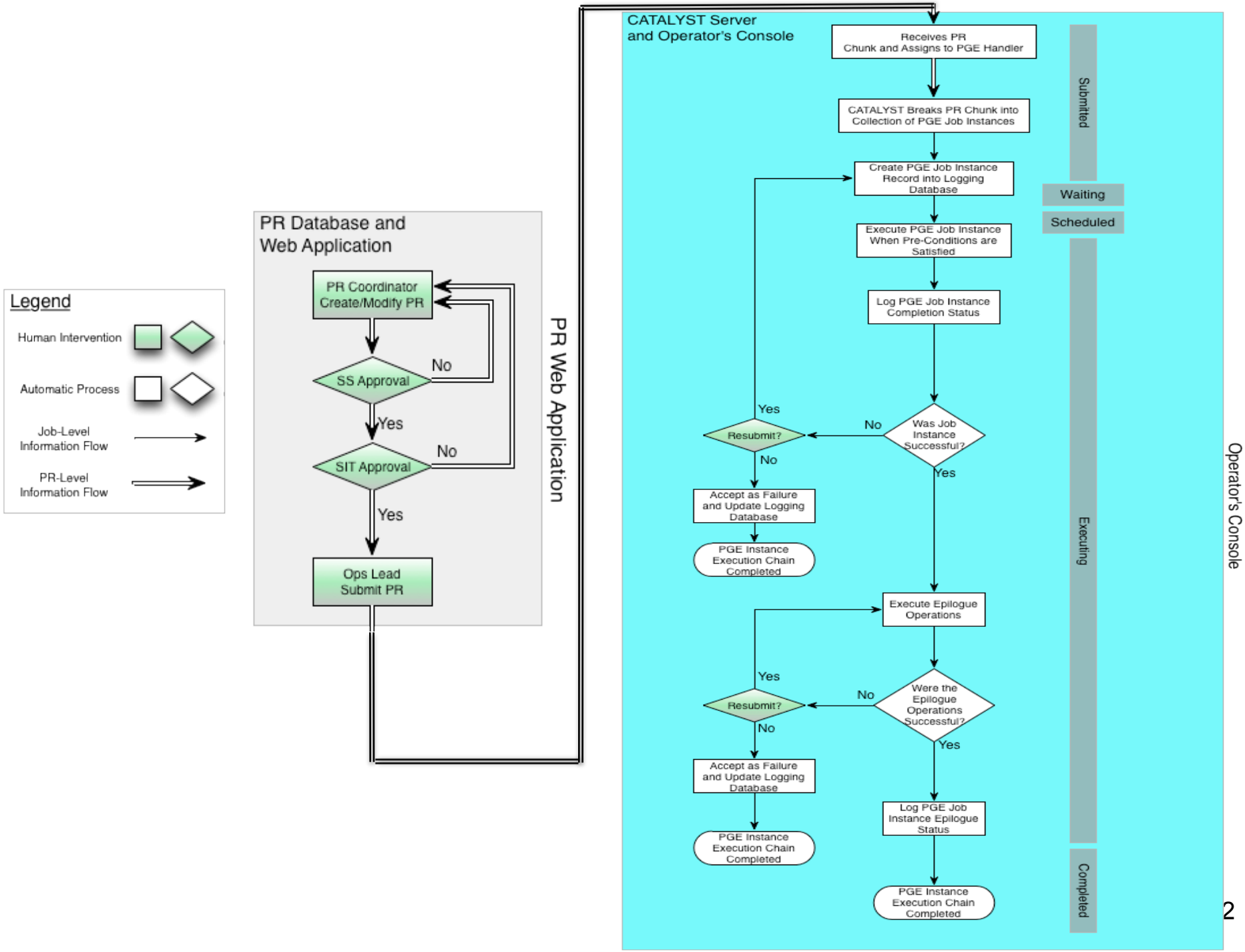
Data Sources

Type of Data	Source	Frequency / # of Files	Volume	Comments
NPP RDR	Land PEATE	131 – 134/day	186MB/day	In case of NPP, RDRs also contains spacecraft diary
VIIRS Sub-sampled Radiance and Geolocation	Land PEATE	288/day	60GB/day	CERES provided code to sub-sample at Land PEATE
VIIRS VAOT	Land PEATE	288/day	840MB/day	Produced by Land PEATE
SURFMAP (Snow/Ice)	NCAR	1/day	66 MB/mo	Used for Clouds and ERBElike Processing
SMOBA Ozone	NCEP	1/day	71MB/mo	
SURFMAP (Snow/Ice)	NOAA/NESDIS	1/day	630 MB/mo	Used for Clouds
MCIDAS Geostationary Data	University of Wisconsin	24/day/satellite	71GB/mo	5 geostationary satellites used per month
MATCH Aerosols	NCAR	2/day	164MB/mo	
Meteorological and Ozone data	GMAO	174 GB/mo	9/Day (CERESG5.2) 8/Day (CERESG5.4)	Currently receiving 2 streams G5.2 and G5.4

CERES Subsystems

- CERES is made up of 7 Working Groups
  - Instrument
  - ERBElike
  - Clouds
  - Inversion or ADM
  - SOFA
  - SARB
  - TISA
- Code organized into 12 Subsystems
  - Each subsystem tied to 1 or more working groups
- Each Subsystem made up of 1 or more Product Generation Executives (PGEs)
  - Currently there are 73 active PGEs

Production Optimization
via CERES AuTomAted job Loading sYSTem (CATALYST) and the Production Request Database


Legend
Human Intervention
Automatic Process
Job-Level Information Flow
PR-Level Information Flow
PR Database and Web Application
PR Coordinator Create/Modify PR
SS Approval
No
Yes
SIT Approval
No
Yes
Ops Lead Submit PR
PR Web Application
CATALYST Server and Operator's Console
Receives PR Chunk and Assigns to PGE Handler
CATALYST Breaks PR Chunk into Collection of PGE Job Instances
Create PGE Job Instance Record into Logging Database
Execute PGE Job Instance When Pre-Conditions are Satisfied
Log PGE Job Instance Completion Status
Was Job Instance Successful?
No
Yes
Resubmit?
Yes
No
Accept as Failure and Update Logging Database
PGE Instance Execution Chain Completed
Execute Epilogue Operations
Were the Epilogue Operations Successful?
No
Yes
Resubmit?
Yes
No
Accept as Failure and Update Logging Database
PGE Instance Execution Chain Completed
Log PGE Job Instance Epilogue Status
PGE Instance Execution Chain Completed
Submitted
Waiting
Scheduled
Executing
Completed
Operator's Console

Production Request Database

- Proposed in Senior Review process to streamline management of Production Requests (PRs): small LOE for multiple years due to funding
- Purpose: Transition PRs from paper to database format to simplify creation and approval process.
- Leverage PR database and provide interface for operators to "submit" a given PR to run on production system & provide searchable job exit status logging function
- Web Interface functionality:
  - Create (developed and tested)
  - Subsystem approval (developed and tested)
  - SIT approval (developed, testing in progress)
  - Operator submit & track (development & testing in progress)
- End-to-End testing May 22nd – June 4th
- Go-Live (use database and interface for all PR management) July 11th

CATALYST

- Been in “pre-design” phase for last 9 months – Significant activity started post NPP launch.
- Primary Functions:
  - Accept PR and build all individual job instances
  - Monitor for dependency / precursor satisfaction
  - Submit job to Sun Grid Engine when ready
  - Throttle job submissions
  - Log job results
- Implements PGE specific logic in individual modules for each PGE. (easy to add, remove or change PGEs)
- Client/Server software model (Server on AMI-P head node, client on operator’s local machine)

Basic "going in" requirements (CATALYST)

- Maximize AMI-P system throughput by partially automating job submission.
- Leverage existing software/hardware infrastructure: AMI Job Submission Scripts (AJSS), AMI-P, PR Database, etc.
- Scalable design with ability to add/modify PGEs
- Manual fallback feature for operators – can call existing AJSS manually to submit jobs
- Transparent and convenient as possible for subsystem developers